

# Cartographier la science et la technologie

jeux d'échelles et dimensions  
d'analyse

Trois fils conducteurs :

- Parcourir les projets de recherche par les **visualisations**
- Classer les projets par **échelles et types d'analyses**
- Sous l'angle des **méthodes**

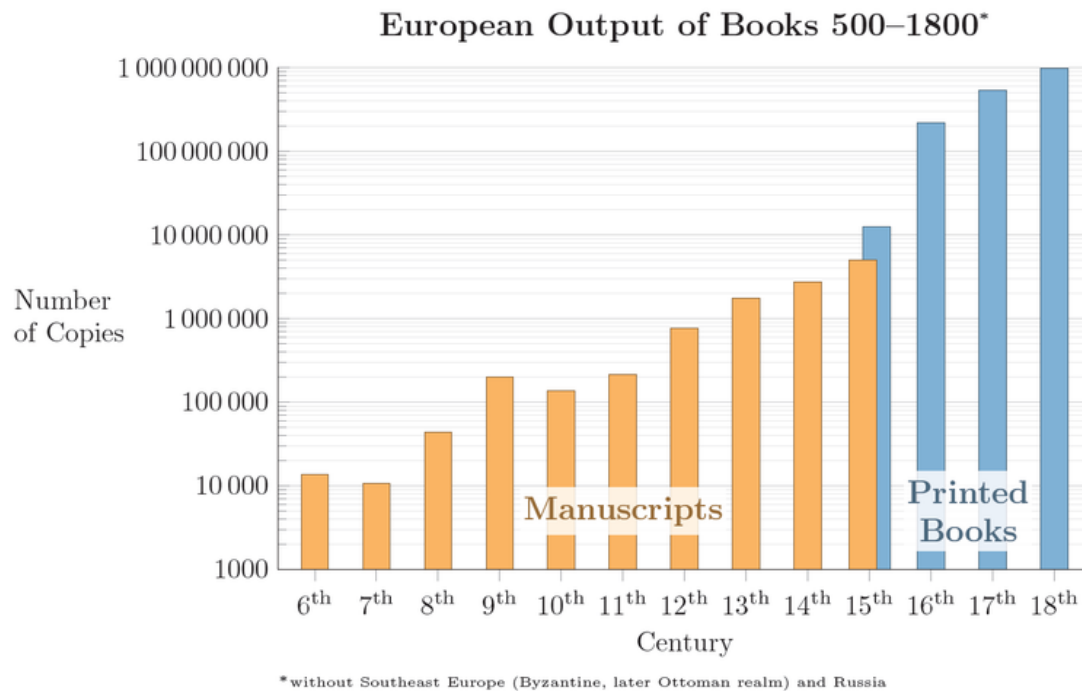
Présentation en **trois parties** :

- **Contexte**, enjeux pour les SHS et cartographie de l'information en STS
- Réflexions autour des **échelles** et des **types d'analyses**
- **Présentation de projets** de recherche et de cartographies

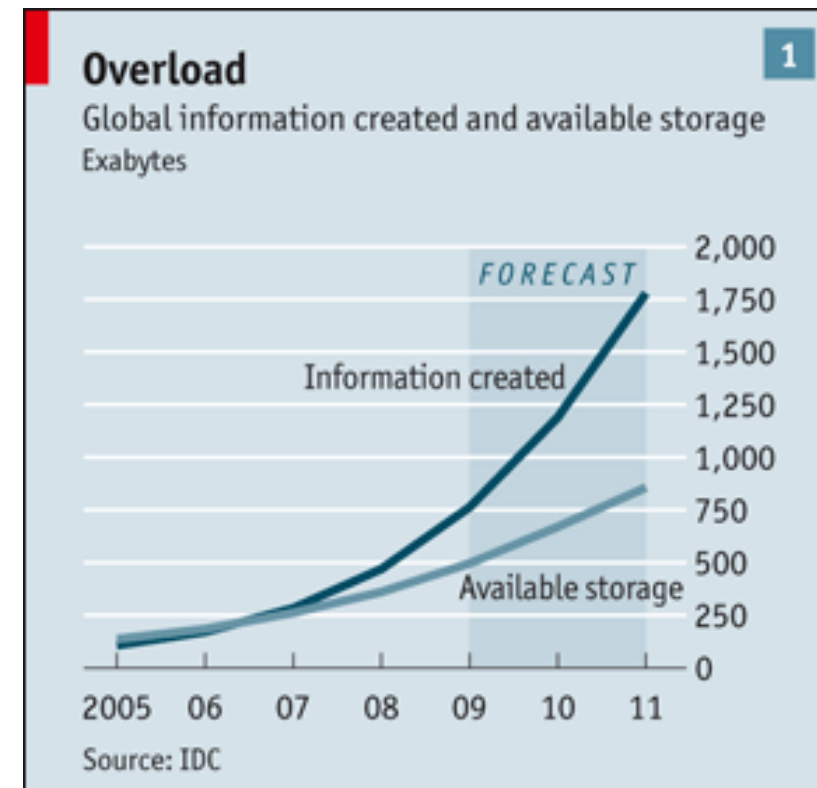
C'est donc avant tout une présentation d'**usages particuliers** de cartes en STS (analyse de la science et de la technologie)

# Production d'information

Les activités humaines, la **vie en société**, **génèrent des données et des informations**. Plus les sociétés se développent, plus le volume créé est important.



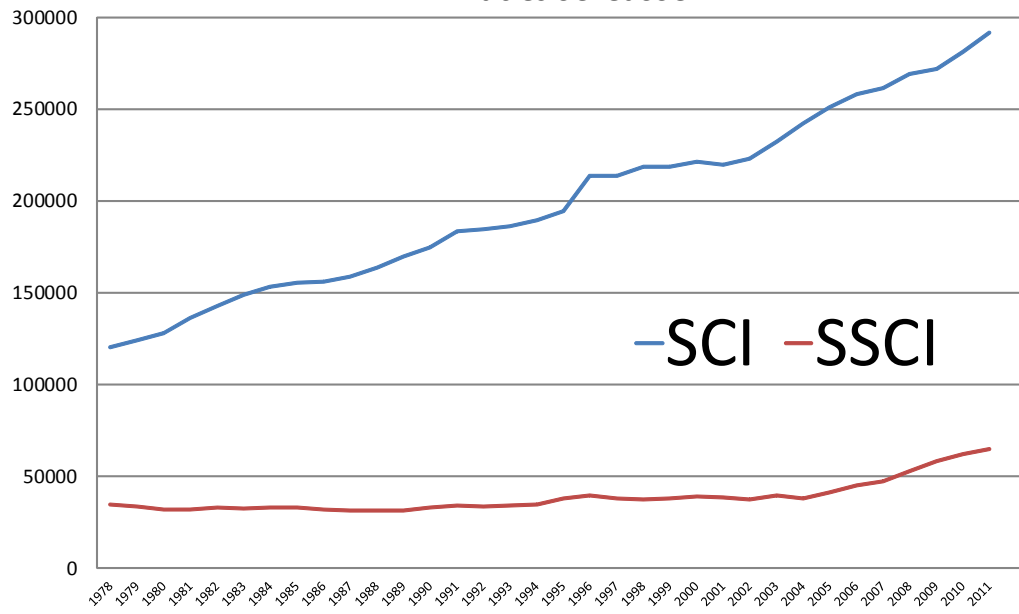
*History of books*, Wikipedia, visité le 06/02/2010,  
[http://en.wikipedia.org/wiki/History\\_of\\_books](http://en.wikipedia.org/wiki/History_of_books)



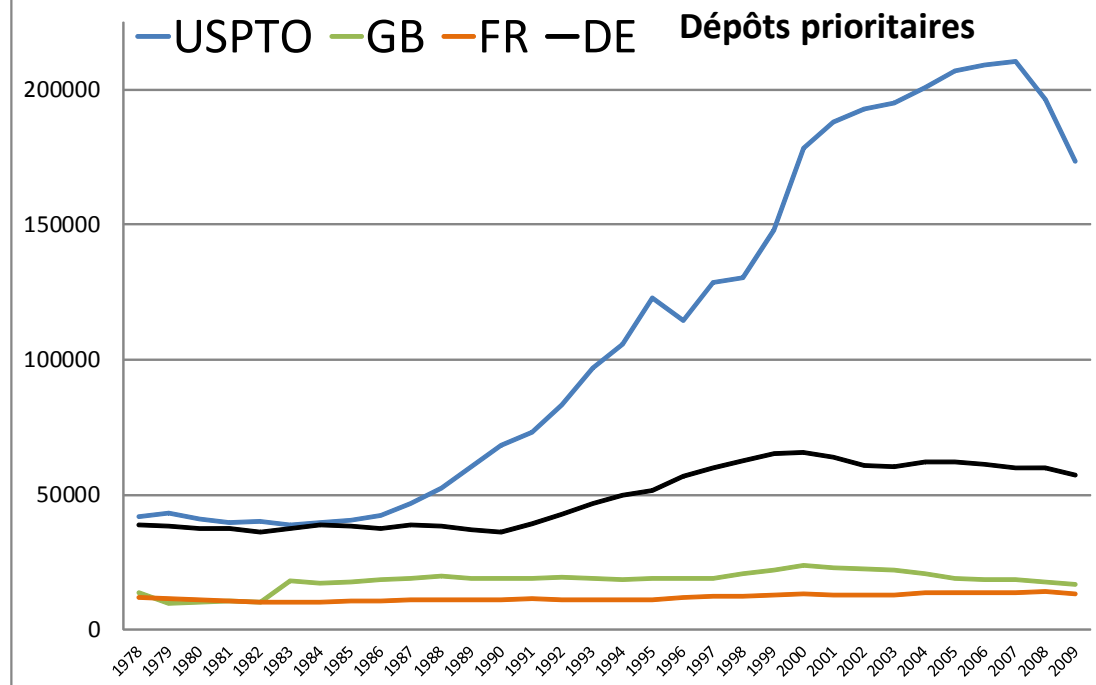
*Data, data everywhere*, The economist, 25/02/2010,  
[www.economist.com/node/15557443](http://www.economist.com/node/15557443)

# Croissance de l'information scientifique et technologique produite

**Publications scientifiques aux Etats-Unis**  
Articles SCI et SSCI



**Brevets**



Données numériques produites dans **les différentes aires** d'activités humaines :

- Travail
- Politique
- Loisir
- ...

## Rencontre avec l'utilisateur

Convergence de facteurs :

- **Accessibilité** accrue : Internet, téléphonie mobile...
- Augmentation des moyens de **traitements** : calcul, stockage, démocratisation des outils et méthodes

Mais l'information n'est pas connaissance.

**Plus le volume de données auquel nous avons accès augmente, plus il est accessible et plus il est difficile de les trier, de l'organiser, et finalement de comprendre l'information véhiculée par ces données.**



*Crunching the numbers, The economist, 19/05/2012, [www.economist.com/node/21554743](http://www.economist.com/node/21554743)*

Nouveaux **défis pour les sciences humaines** et sociales :

- **Nouveaux objets** : variété des objets manipulés (textes, enregistrements audio, vidéo...) et des relations entre objets (hypertext...)
- Nouvelles **méthodes** et **capacités** de fouilles

## Pourquoi cartographier l'information ?

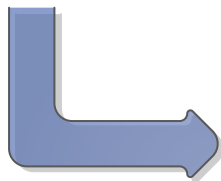
Les données sont parfois si complexes que **le meilleur moyen de les expliquer est d'utiliser l'image.**

« ... Il faut **construire [la visualisation] et l'organiser jusqu'à ce que toutes les relations entre les données en soient révélées.** »

Jacques Bertin, *La graphique et le traitement graphique de l'information*, Flammarion, 1977

Avec un diagramme, une carte, un histogramme ...

- Chaque élément est défini à l'avance
- Chaque signe représenté a une seule signification (monosémie)
- Le travail de lecture se situe **entre les significations**



Le lecteur s'interroge alors sur **les relations** : quelles sont les relations qui s'établissent entre tous les signes, entre tous les éléments représentés ?

# Et la cartographie de la science et de la technologie ?

A la confluence de :

- **l'analyse de la science et de la technologie** : comme l'étude du développement d'un champs scientifique et/ou technologique. Les données analysées sont donc le plus souvent les publications scientifiques et les brevets.
- **la visualisation de l'information** avec un nombre important de vecteurs possibles : carte géographique, graphique, réseaux d'acteurs, cartes sémantiques ...

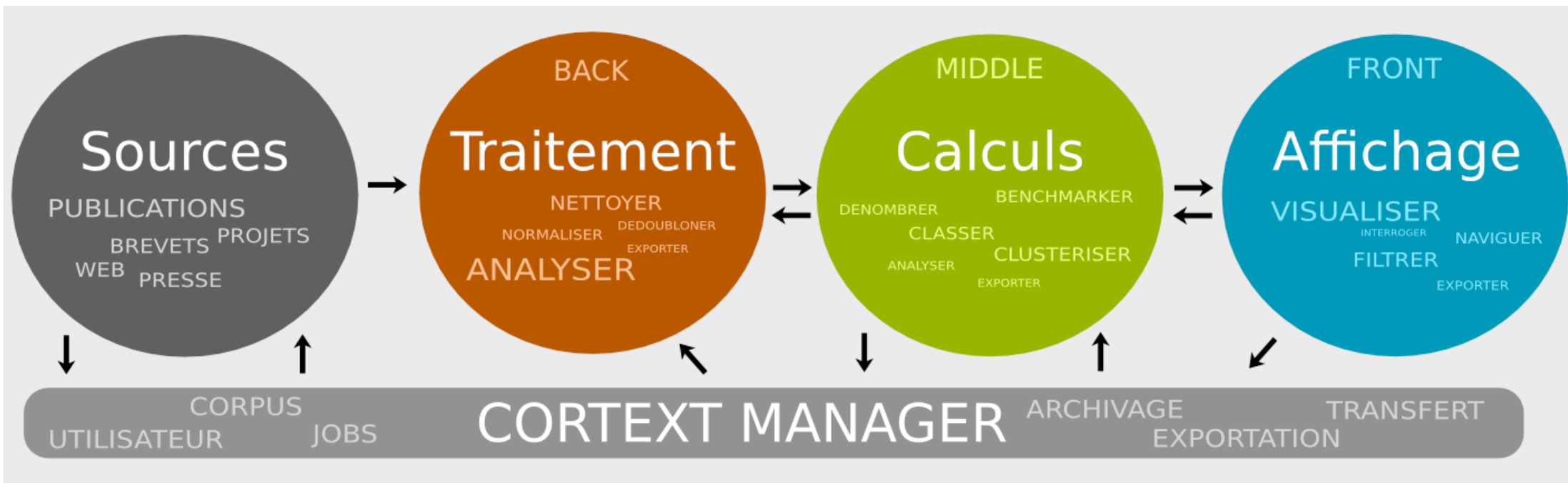


## Mais pourquoi cela est-il intéressant ?

Il est alors possible de comprendre visuellement, plus immédiatement, des situations, des problèmes souvent complexes.

- **Synthèse** : présentation synthétique de différentes dimensions d'un problème, d'un sujet
- **Positionner** : la position des thématiques, des acteurs, ...
- **Vecteur performant** : permet au lecteur (scientifique, citoyen, politique) de comprendre les relations qui associent les éléments représentés

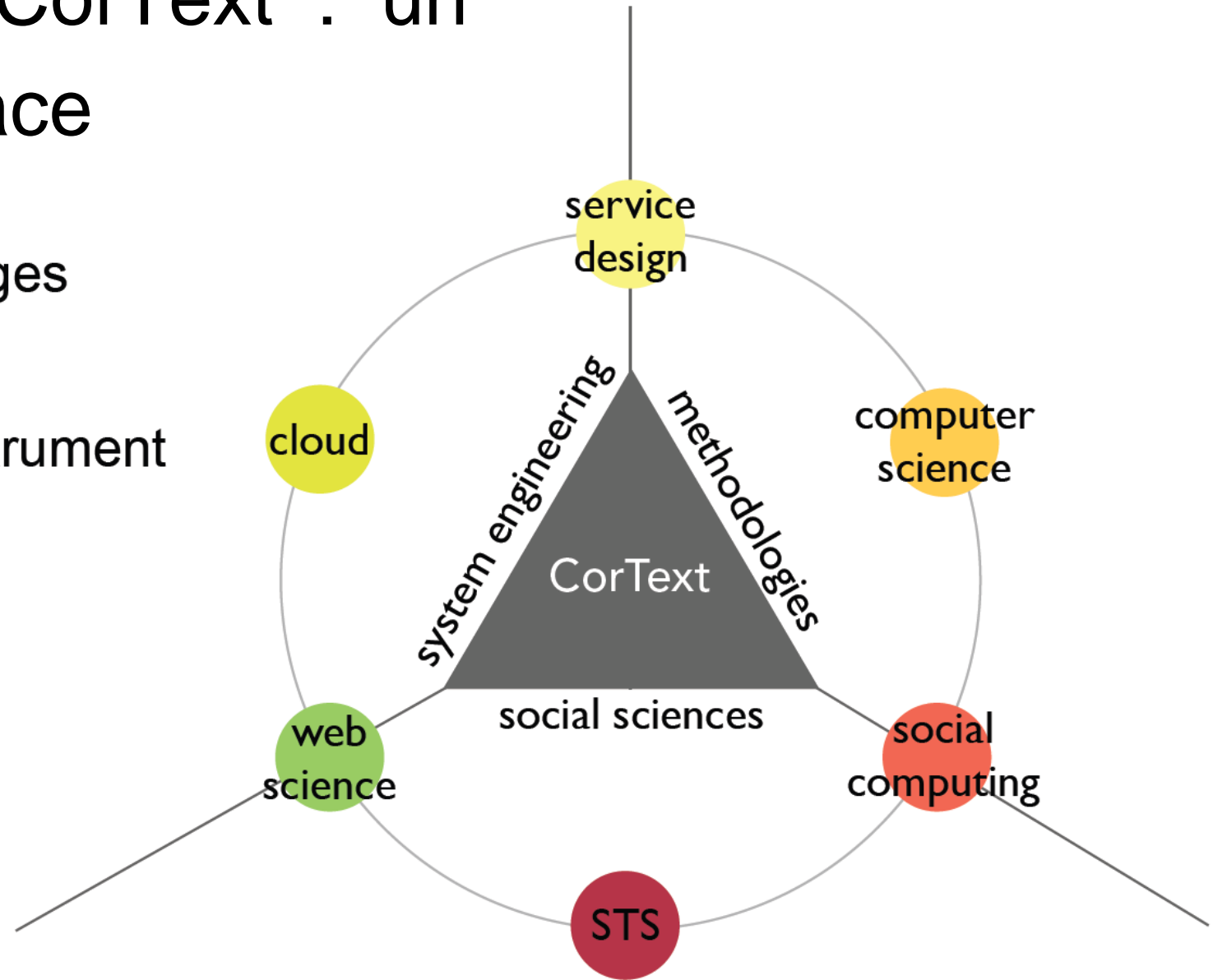
## Des sources aux visualisations



Egalement la délinéation (STS : le champ)

# La plateforme CorText : un outil, une interface

- Scientific challenges
- An integrated instrument
- Project examples



# Echelles et dimensions d'analyse

- Organiser les données
- Échelles d'analyse
- Proposition de grille pour organiser les projets

# Organiser les données

*Information Graphics*, Sandra Rendgen, 2012, Taschen, 480 p.

**Lieu** : les éléments sont organisés de façon spatiale (ex : carte géographique)

- Les évènements se déroulent à un certain endroit
- Les objets sont organisés de façon spatiale, il est donc possible de placer des évènements simultanés dans un espace.

**Temps** : les éléments sont organisés de façon chronologique (ex : graphique d'une variable par année).

- Il y a une séquence temporelle fixe, chronologique
- Permet de décrire un phénomène, une tendance (mais n'explique pas les causes)

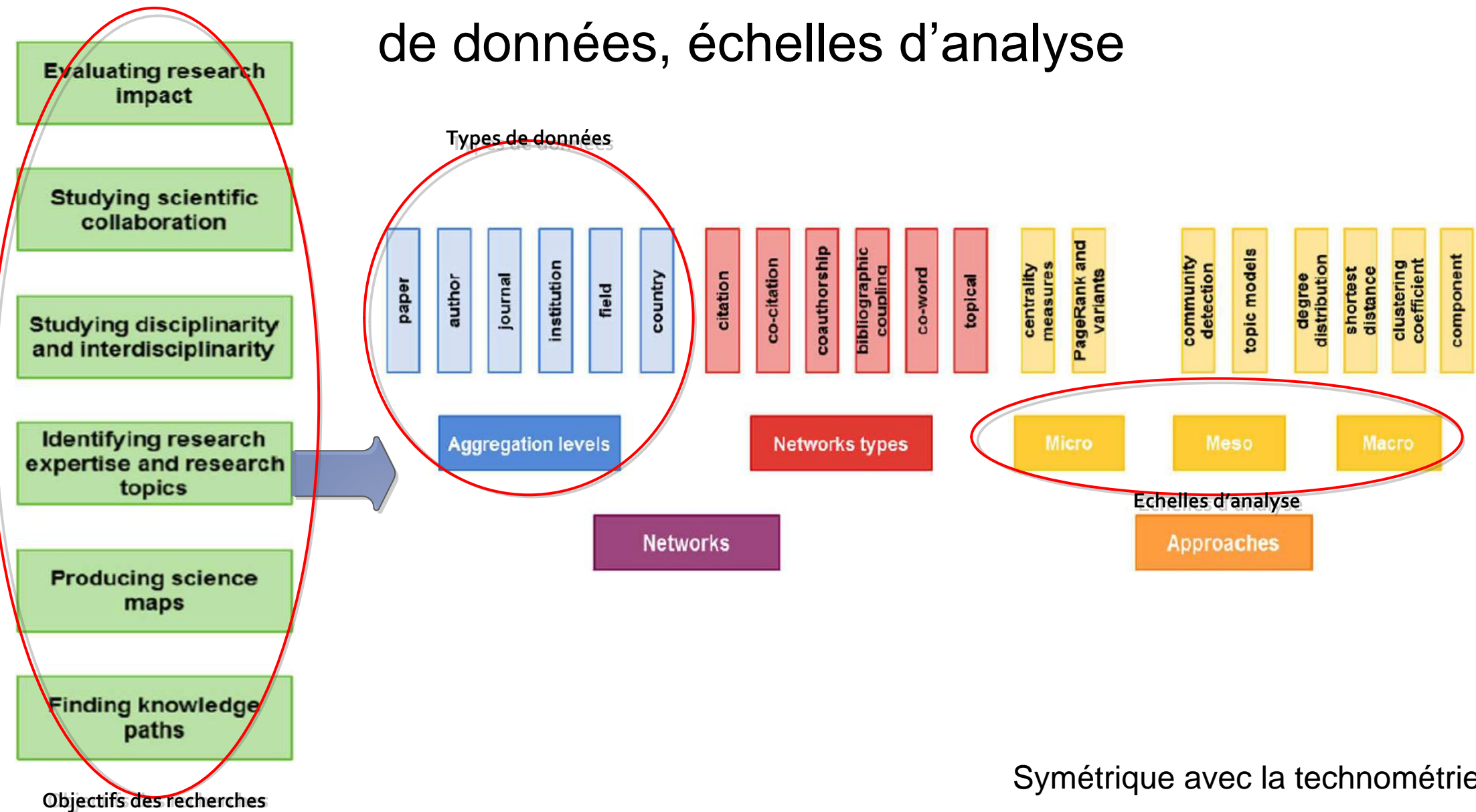
**Catégorie** : les éléments sont divisés en classes (ex : une classification présentant des groupes de données, de phénomènes)

- Par type, par caractéristiques communes
- Il y a un ordre

**Hiérarchique** : les éléments sont organisés verticalement (ex : treemap, camembert...)

- Éléments triés par ordre de priorité, par ordre d'importance
- Présente un classement des éléments, par quantité

# Publications scientifiques : articulation entre objectifs, types de données, échelles d'analyse





Se questionner et interroger les données, les 5 W : When, Where, What, Who, Why ?

Outil, objet, résultat, objectif...

Responsable, acteur, sujet, cible...

Lieu, service...

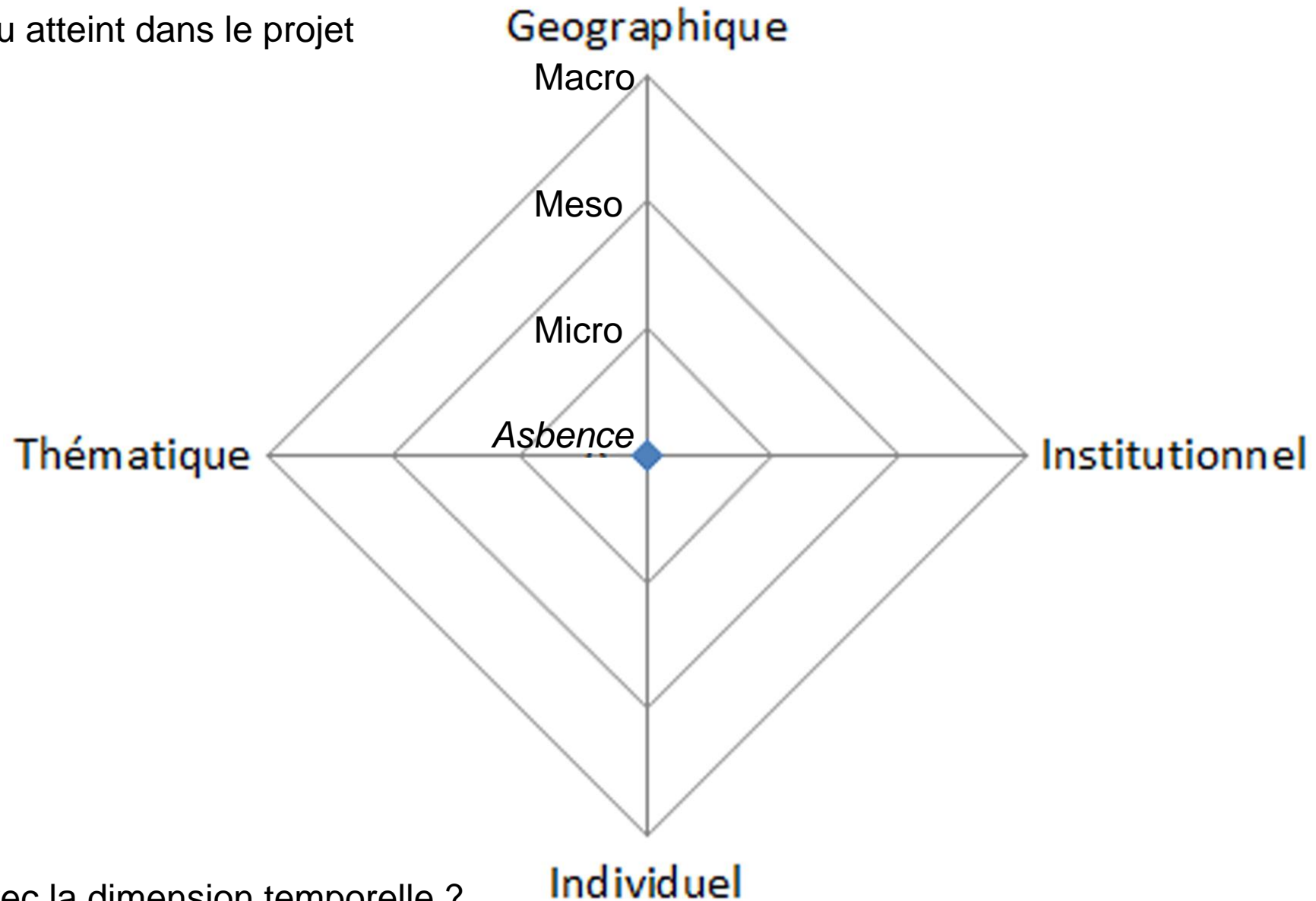
Dates, périodicité, durée...

Echelles par nombre d'enregistrements

	<i>Micro/Individual (1-100 records)</i>	<i>Meso/Local (101-10,000 records)</i>	<i>Macro/Global (10,000 &lt; records)</i>
<b>Statistical Analysis/Profiling</b>	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, SA, all of sci
<b>Temporal Analysis (When)</b>	Funding portfolio of one individual	Research bursts of PNAS	113 Years of P Research
<b>Geospatial Analysis (Where)</b>	Career trajectory of one individual	Mapping a scientific intellectual	PNAS
<b>Topical Analysis (What)</b>			VxOrd/Topic NIH funding
<b>Network Analysis (With Whom?)</b>	NSF work of one		NIH's

## Grilles de lecture des projets :

- Dimensions d'analyse
- Echelles : niveau atteint dans le projet



Variables croisées avec la dimension temporelle ?

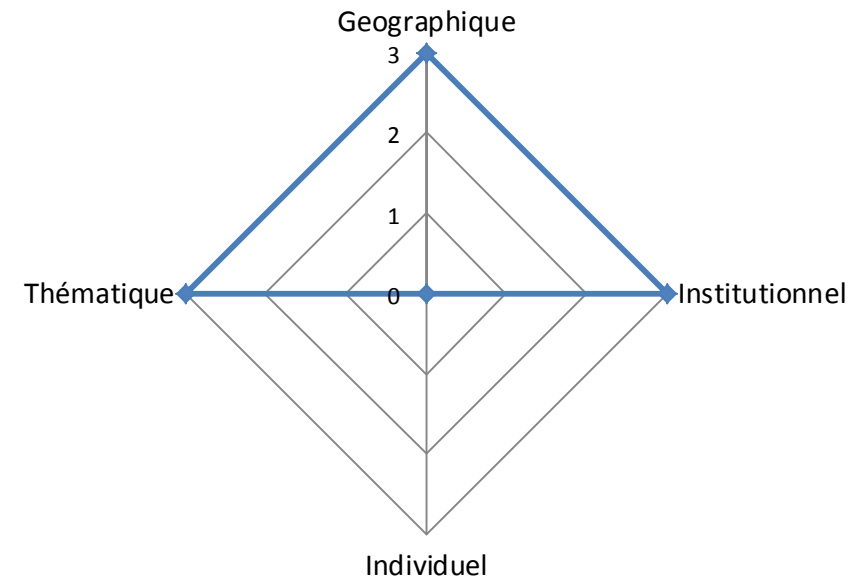


# Projets de recherche et visualisations

- CIB
- Développement des nanotech
- Carte globale de la technologie
- Positionner le nucléaire
- Catalyse hétérogène et co-activité brevets / publications

# Globalisation de la R&D des 2 400 plus grands groupes mondiaux en investissements de R&D

## Corporate Invention Board



### Extraction

- 2 000 groupes du scoreboard (IPTS)
- 400 filiales chinoises et indiennes
- 170 000 filiales (ORBIS)
- 6 millions de brevets prioritaires (Patstat 2009)

### Enrichissements

- Graphies de noms de groupes
- Méthodes visant à compléter les adresses d'inventeurs

### Éléments principaux de méthode

- Normalisation des noms de déposants Tom Magerman (2009) et Bart Van Looy (2010)

## Corporate Invention Board

For an in-depth analysis of global patent portfolios

recherche...

Accueil Méthodologie Sources de données Corporate Rankings Exploitation du CIB Visualisation News Le projet

Accueil > Visualisation

Expand view

ALL (Technology) / ALL (Inventors localisation) / 1985 - 2005

Full Screen

### Date period

- ☐ 1986-1995  
☐ 1996-2005  
☒ 1986-2005

### Selection

#### Spread

Headquarter's location

Industrial activity

Inventor's location

Technology

Company name

#### Over

Headquarter's location

Industrial activity

Inventor's location

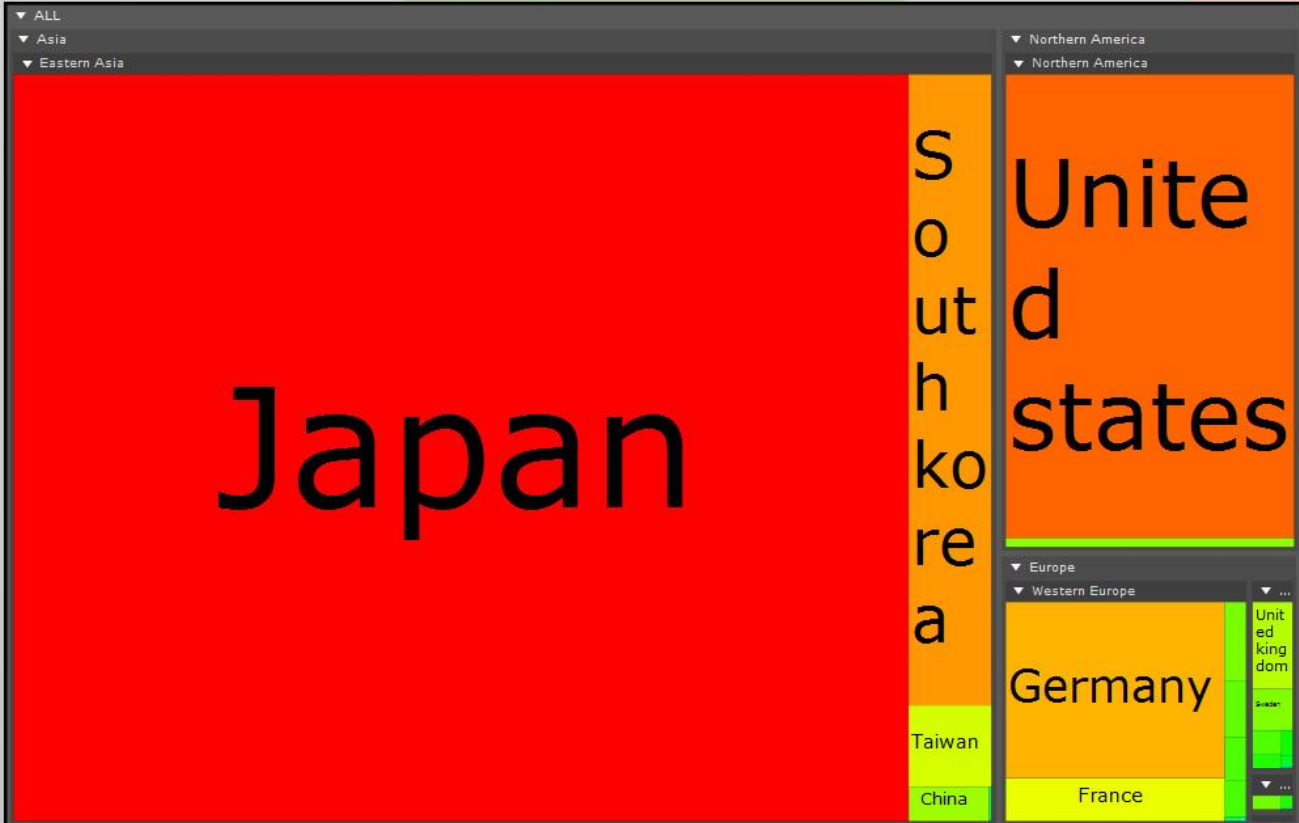
Technology

Company name

Display

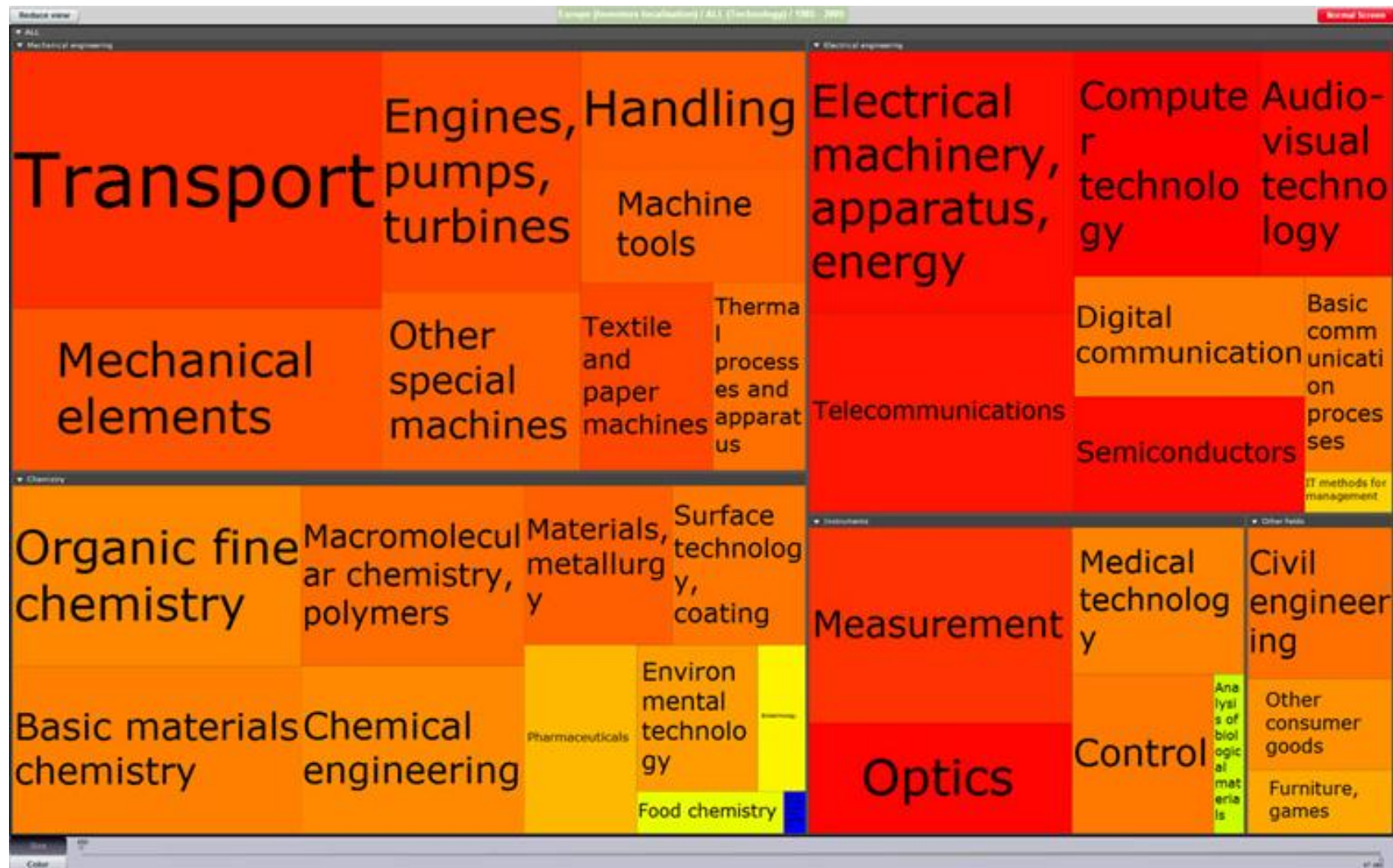
Export

Help



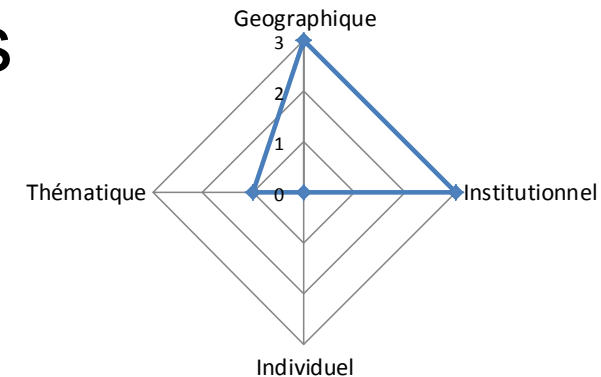
Size

Color



# Etude d'un champ scientifique et technologique en forte croissance : les nanotechnologies

## Nanotechnologies



## Délinéation

- Requête lexicale dynamique par année basée sur l'analyse morphosyntaxique d'un cœur de documents (Andrei Mogoutov, 2007)
- Extractions lexicales par années des publications et des brevets

## Enrichissements

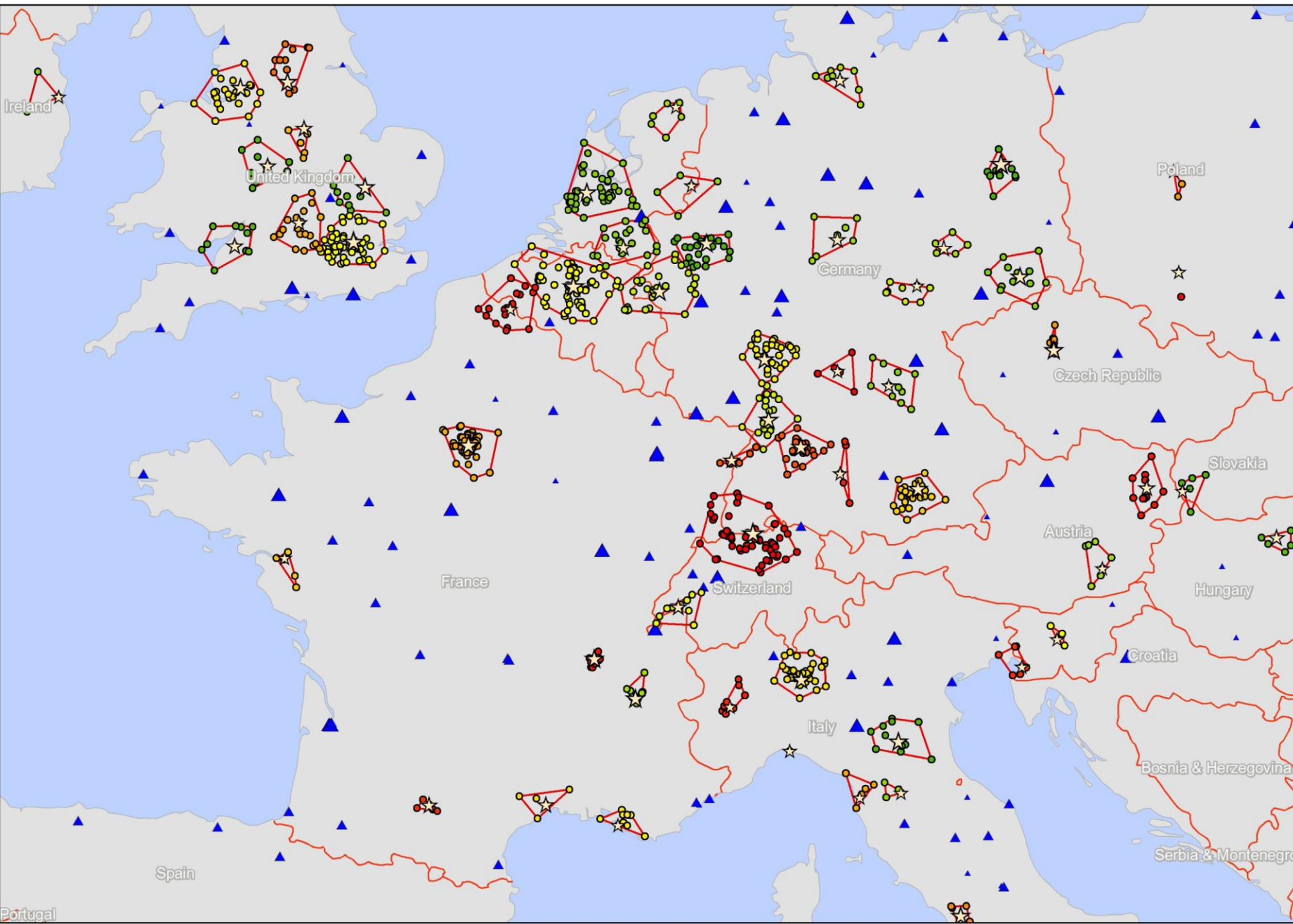
- Localisation et géolocalisation des adresses d'inventeurs et d'auteurs
- Clustering géographique
- Harmonisation et catégorisation des institutions

## Principaux éléments de méthode

- Requête lexicale dynamique par année basée sur l'analyse morphosyntaxique d'un cœur de documents (Andrei Mogoutov, 2007)
- Clusters géographiques : DBScan (M. Ester, 1996) et Chameleon (George Karypis, 1999)



# Clusters géographiques

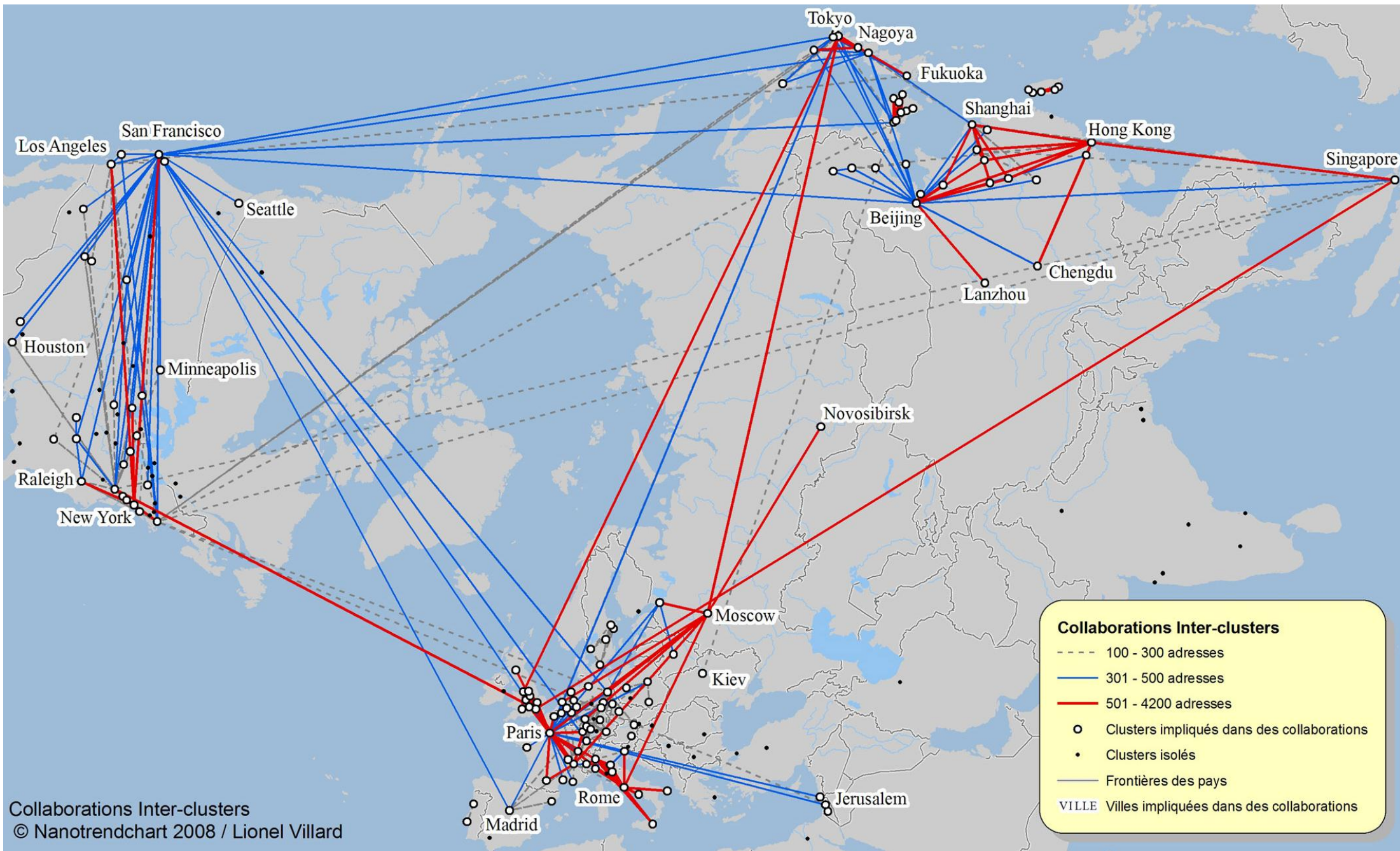


Adresses de  
publications

1998-2006

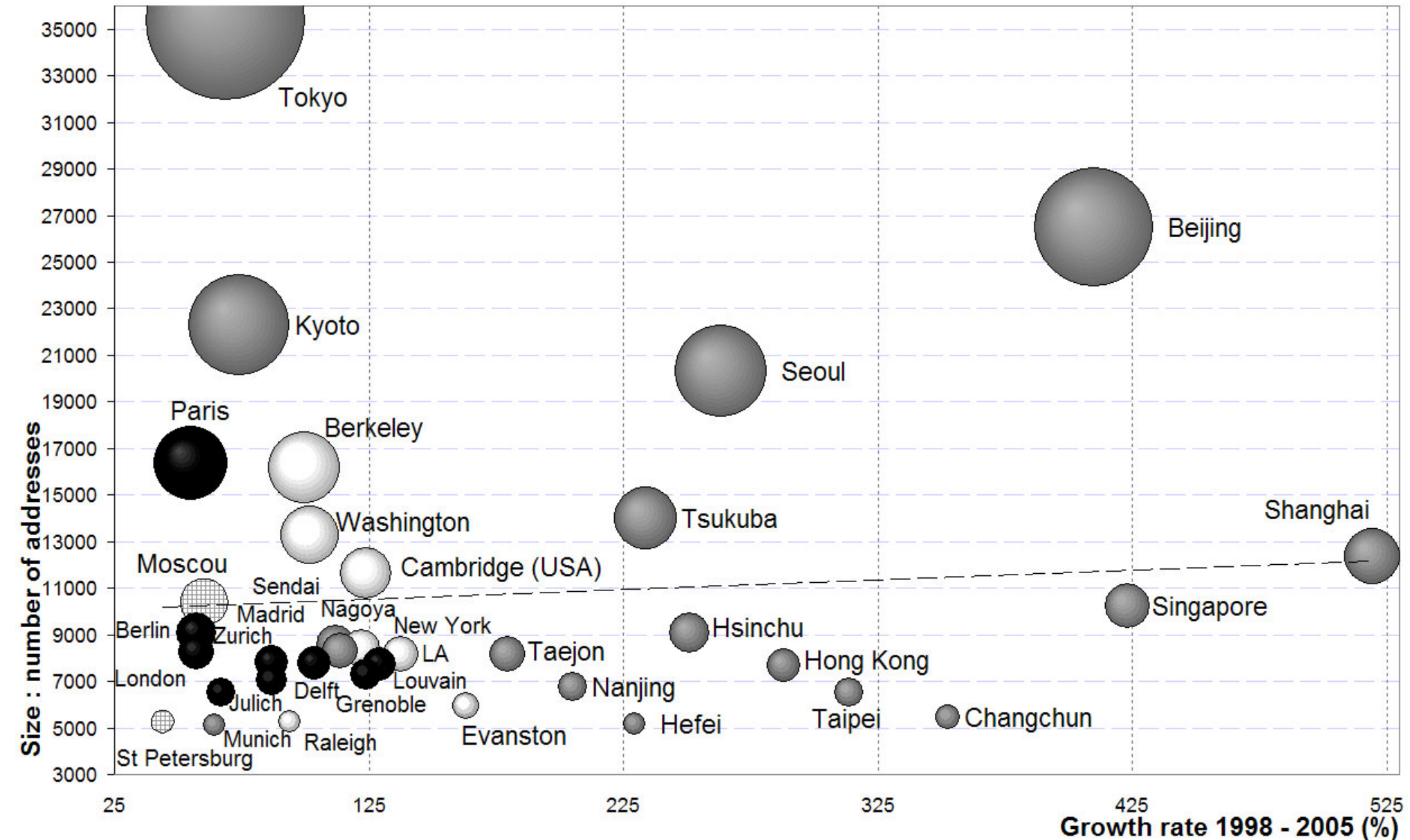
80% de l'activité  
dans 203 lieux





## Big and medium clusters (more than 5 000 addresses)

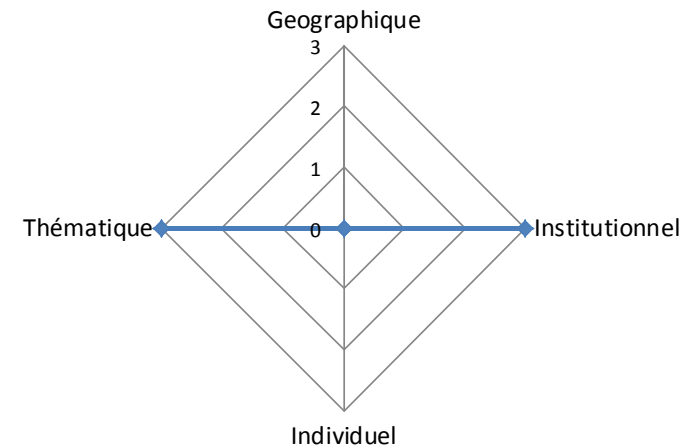
L. Villard, B. Kahane, A. Delemarle, P. Laredo / ESIEE-LATTS-Univ Paris Est / © Nanotrendchart 2009





# Global Map of Technology

Map of Technology

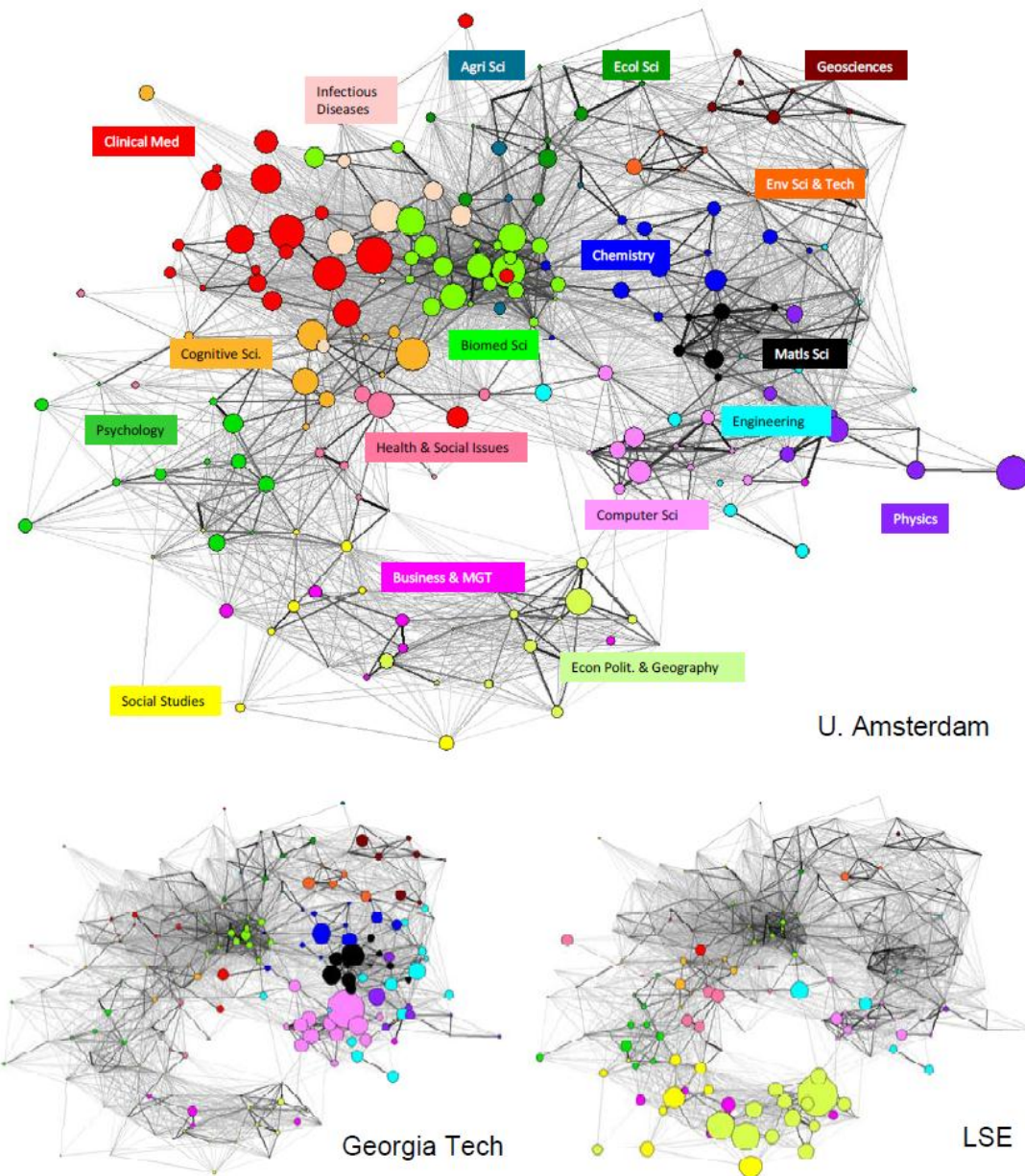


## Périmètre initial

- Ensembles des **dépôts prioritaires** (6 millions) disponibles

## Principaux éléments de méthode

- Classification de technologie produite la WIPO (Ulrich Schmoch, 2008), enrichie : 5 domaines, 35 classes et **389 sous-champs technologiques** basées sur le IPC
- **Co-occurrences d'IPC** et clustering



Motivations ?

## Carte globale de la science:

Publication scientifiques

Citations journals-journals

Catégories ISI

Projections sur cette carte

Figure 3. Publications profiles of the University of Amsterdam, Georgia Tech and London School of Economics (LSE) overlaid on the map of science.



Heating,  
Combusting,  
Iron metallurgy

Coating, Fibers &  
Textile

Chem Phys  
Processes,  
Separation

Compositions Of Macromol Comp

Macromol without C-to-C Unsaturated Bonds

Inorg Or Non-Macromol Organ Subst

Processes Of Compounding

Macromol with C-to-C Unsaturated Bonds

Paints and inks

Materials For Miscellaneous Applications

Macromolecular  
compounds

Medical,  
Cosmetics, Food

Machine  
Tools

Chem and Phys Processes

Printing

Semiconductor Devices

Chem Phys  
Analyses &  
Techno, Biotech

Vehicles,  
Transportation  
engineering

Vehicles

Optical Elements Systems

Semiconductor,  
Optics, Surface  
treatment

Electronics  
Electrical  
engineering

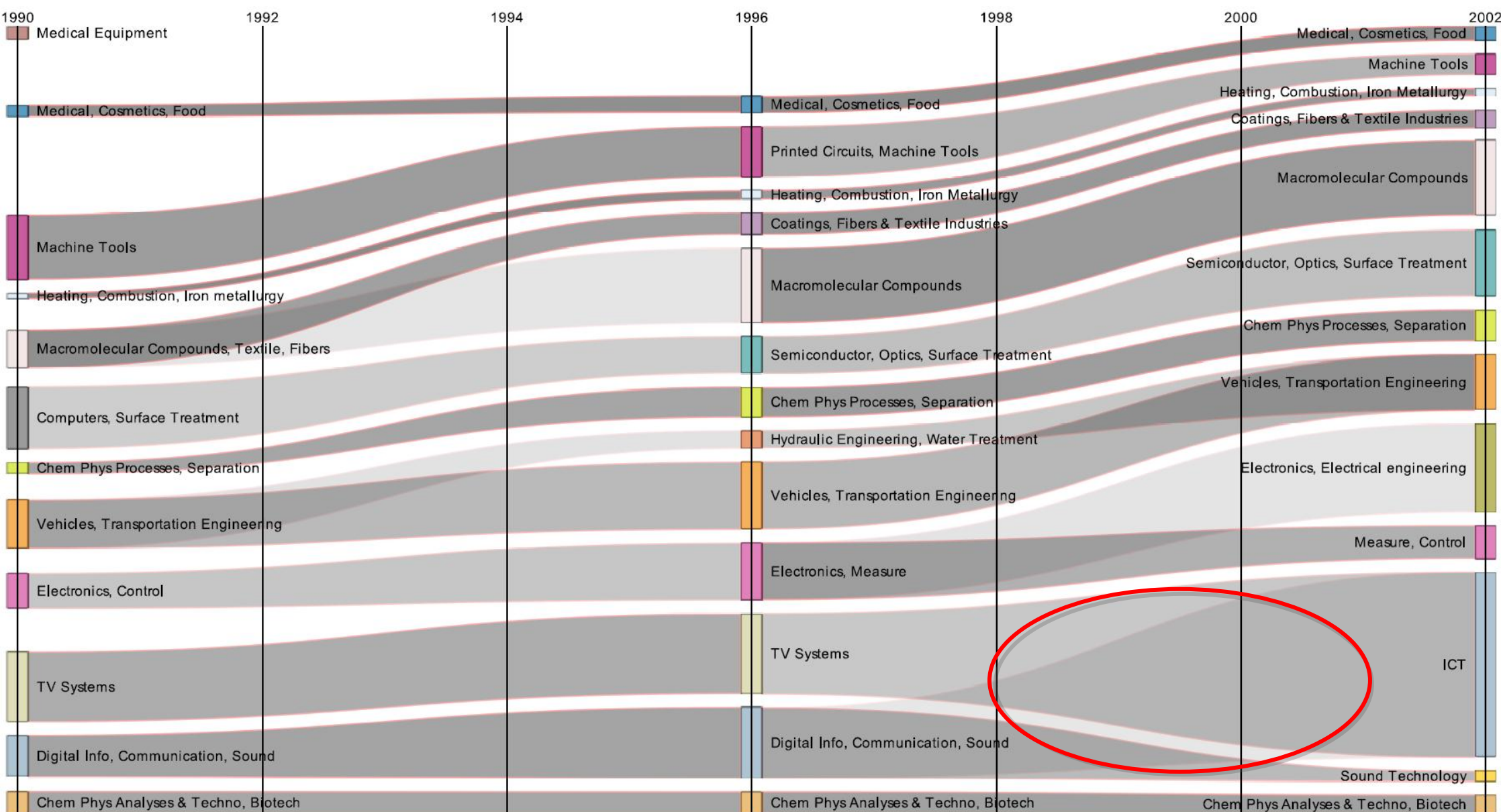
Measure,  
Control

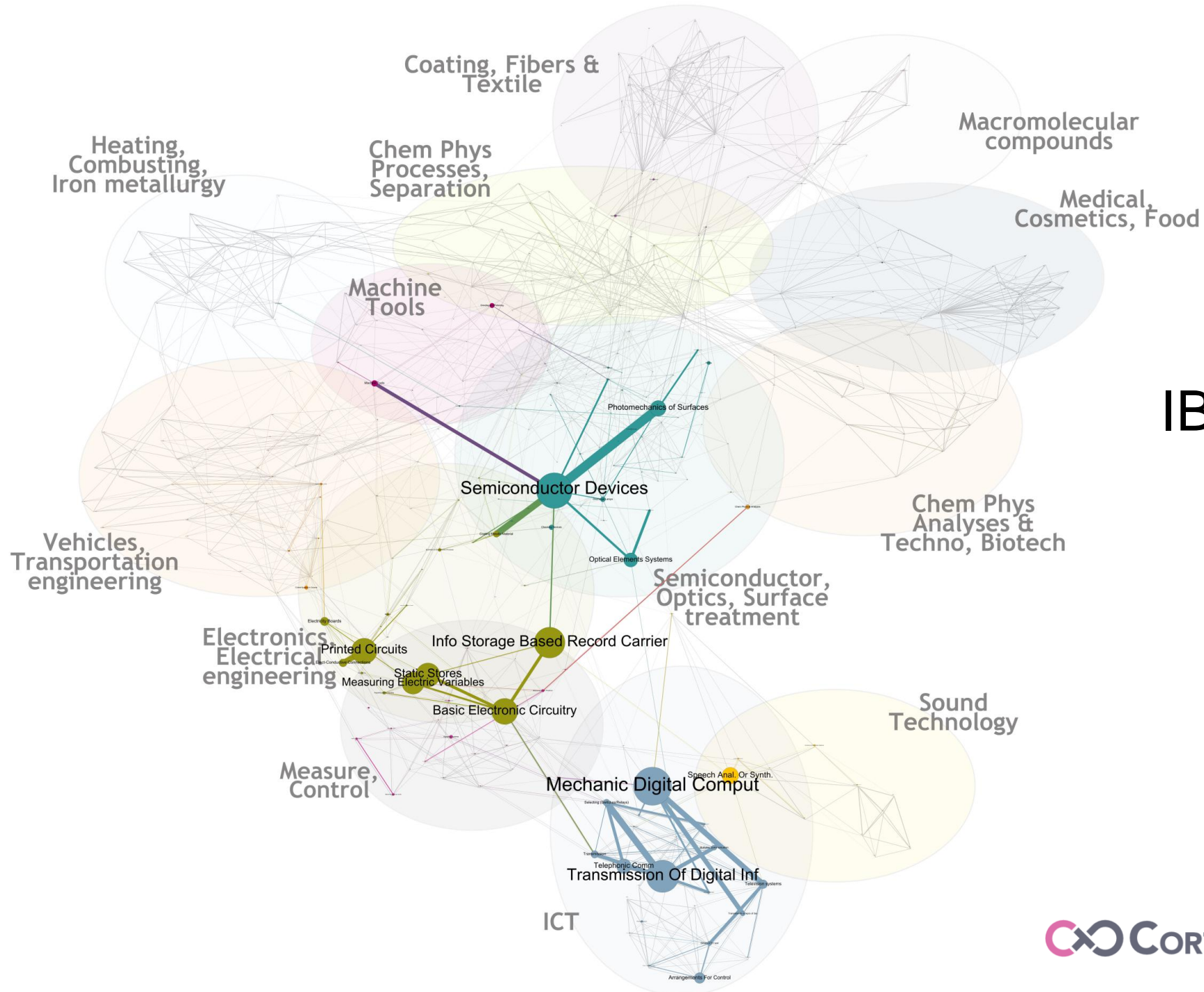
Mechanic Digital Comput

Sound  
Technology

ICT

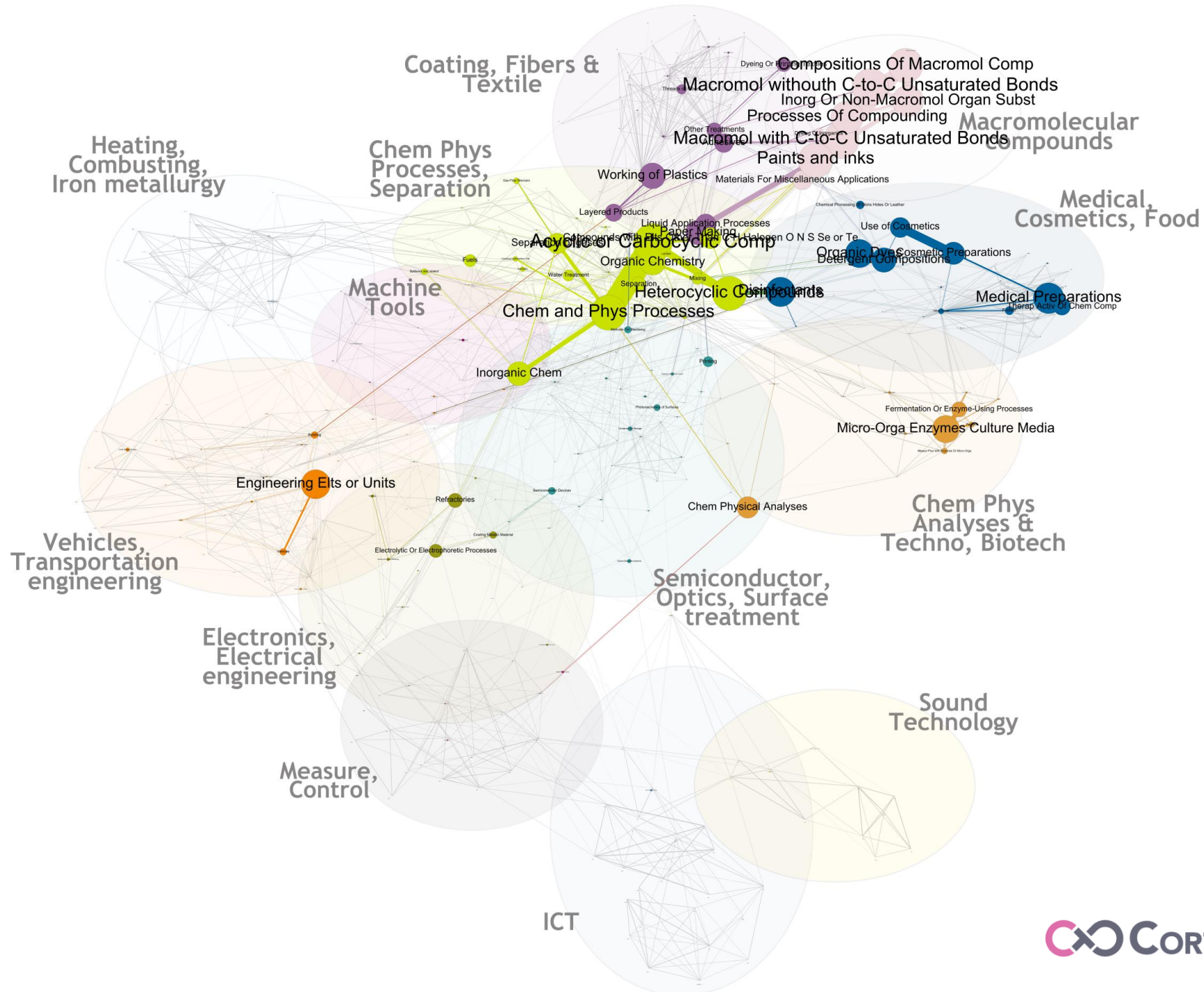






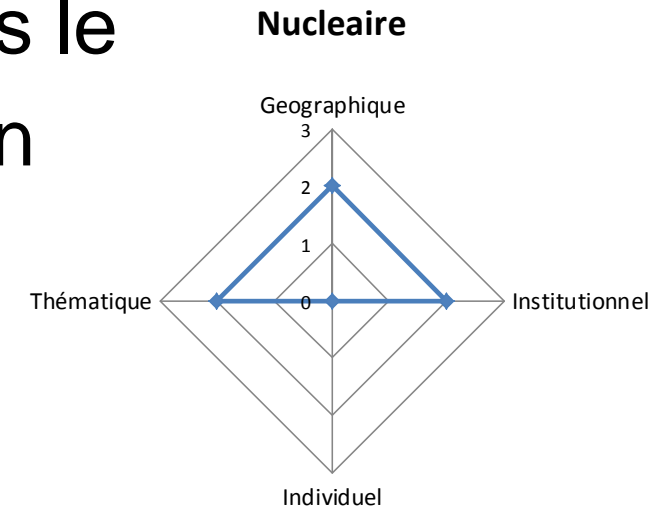
IBM





BASF

# Positionner l'industrie du nucléaire dans le secteur des réservoirs haute pression



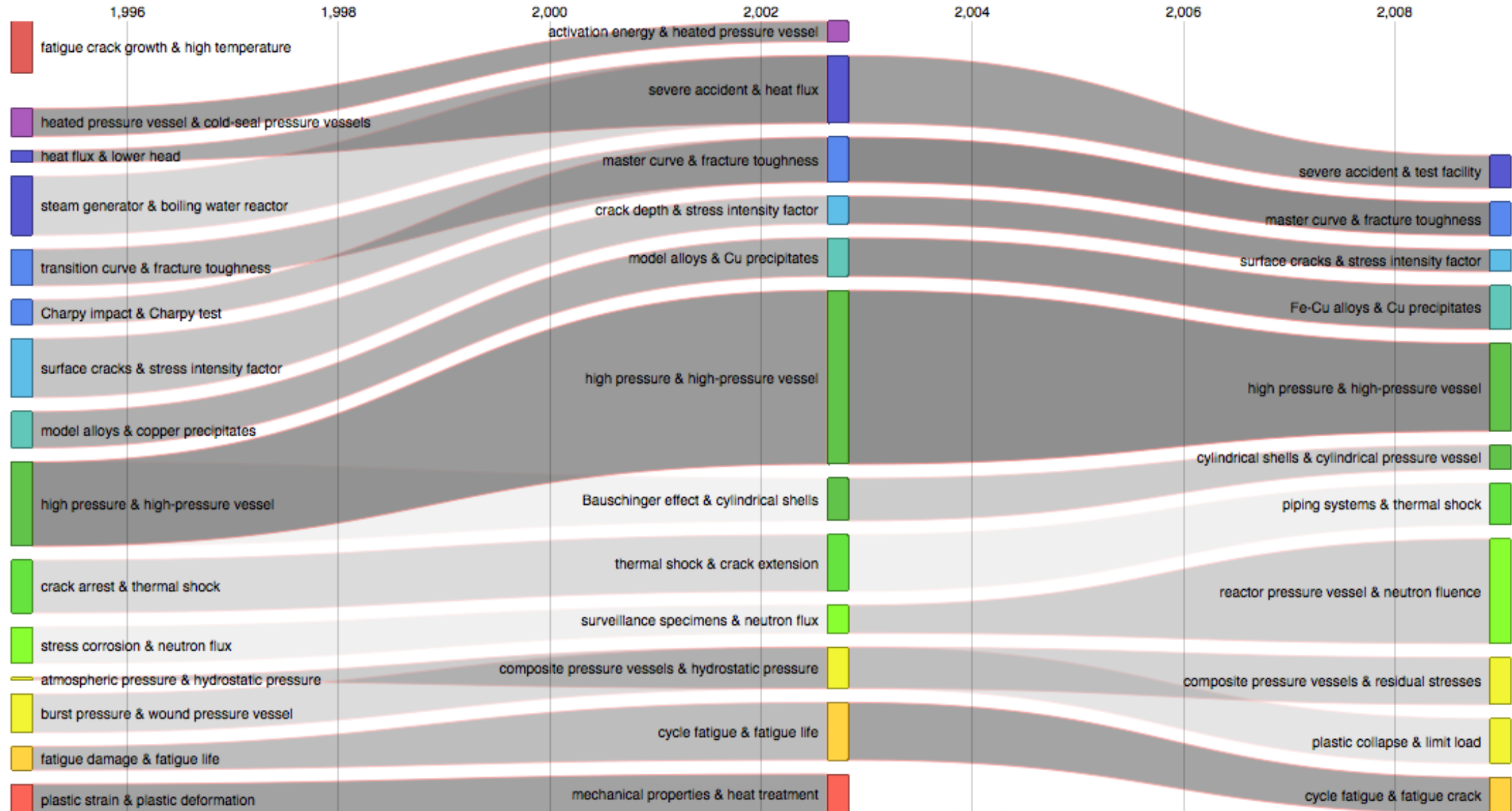
## Délinéation

- Requête lexicale simple

## Éléments principaux de méthode

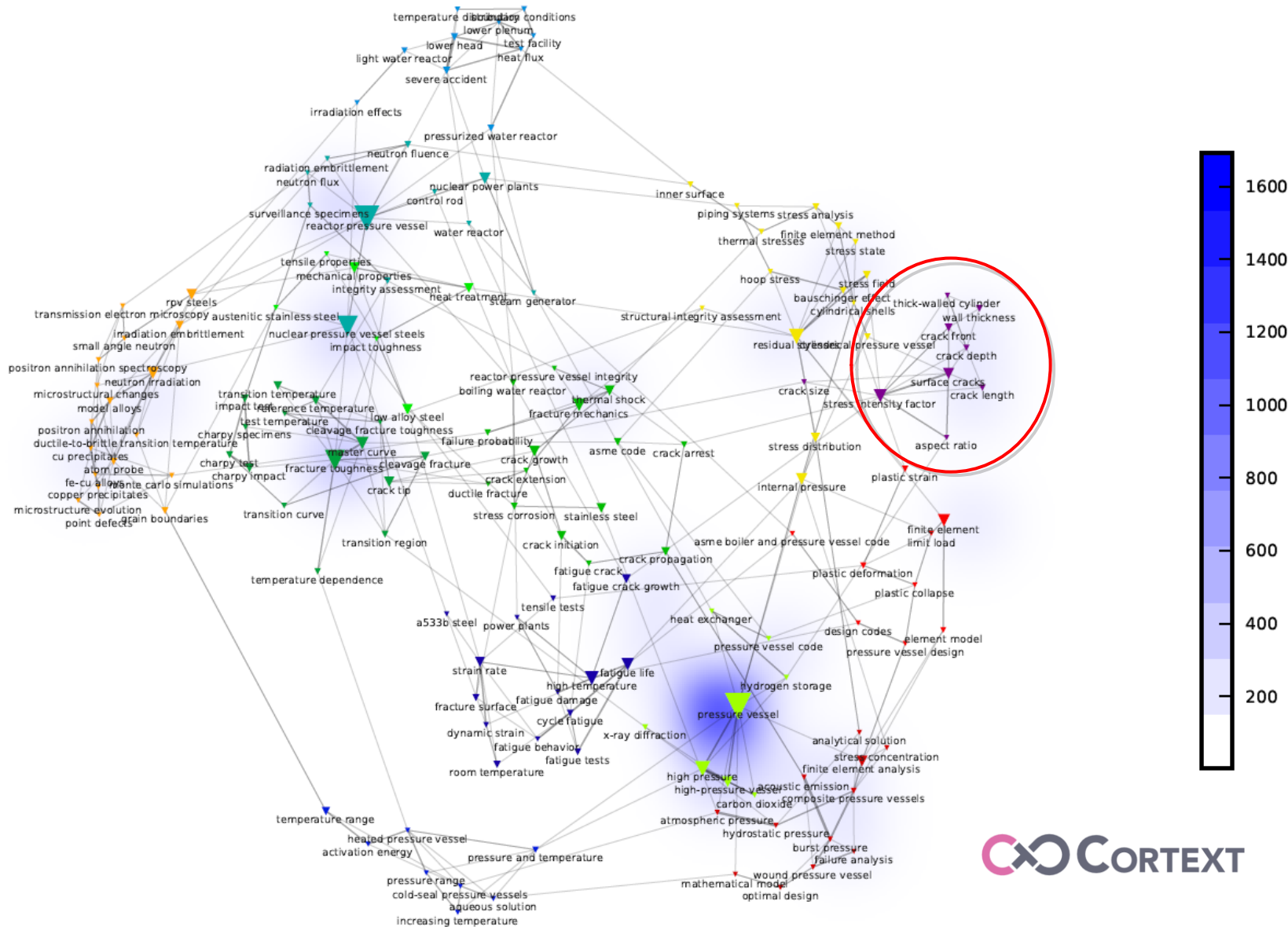
- Analyse morpho-syntaxique, extraction des multitermes
- Projection d'une troisième variable (institutions, pays...) sur le réseau des co-occurrences de termes

# Trois secteurs : imprimerie, cuves haute pression et nucléaire





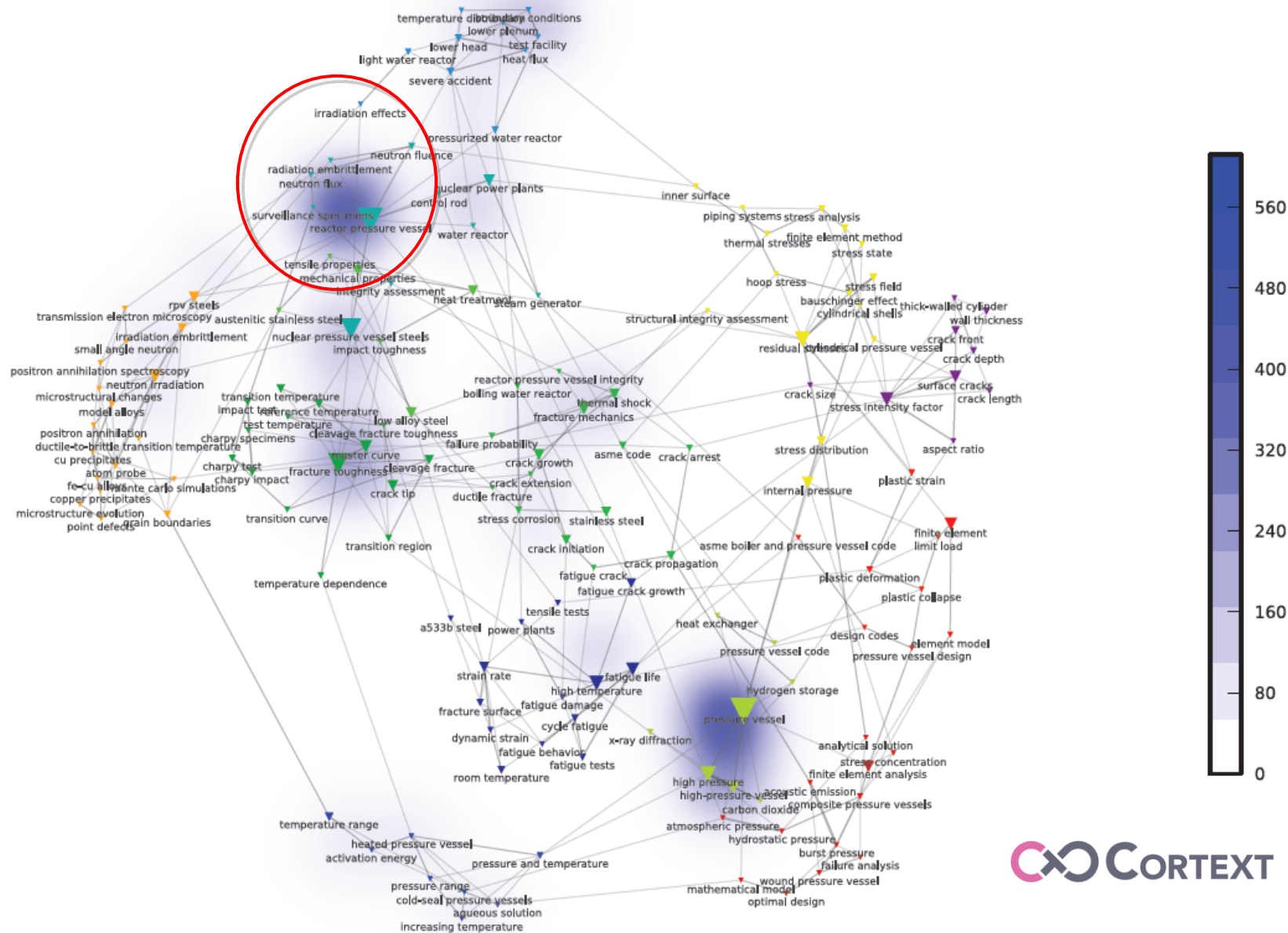
usa, 1991-2012



Publications scientifiques USA,  
00 Pressure vessel, entre 1991 et 2012

Présence sur les questions de résistance.

germany, 1991-2012

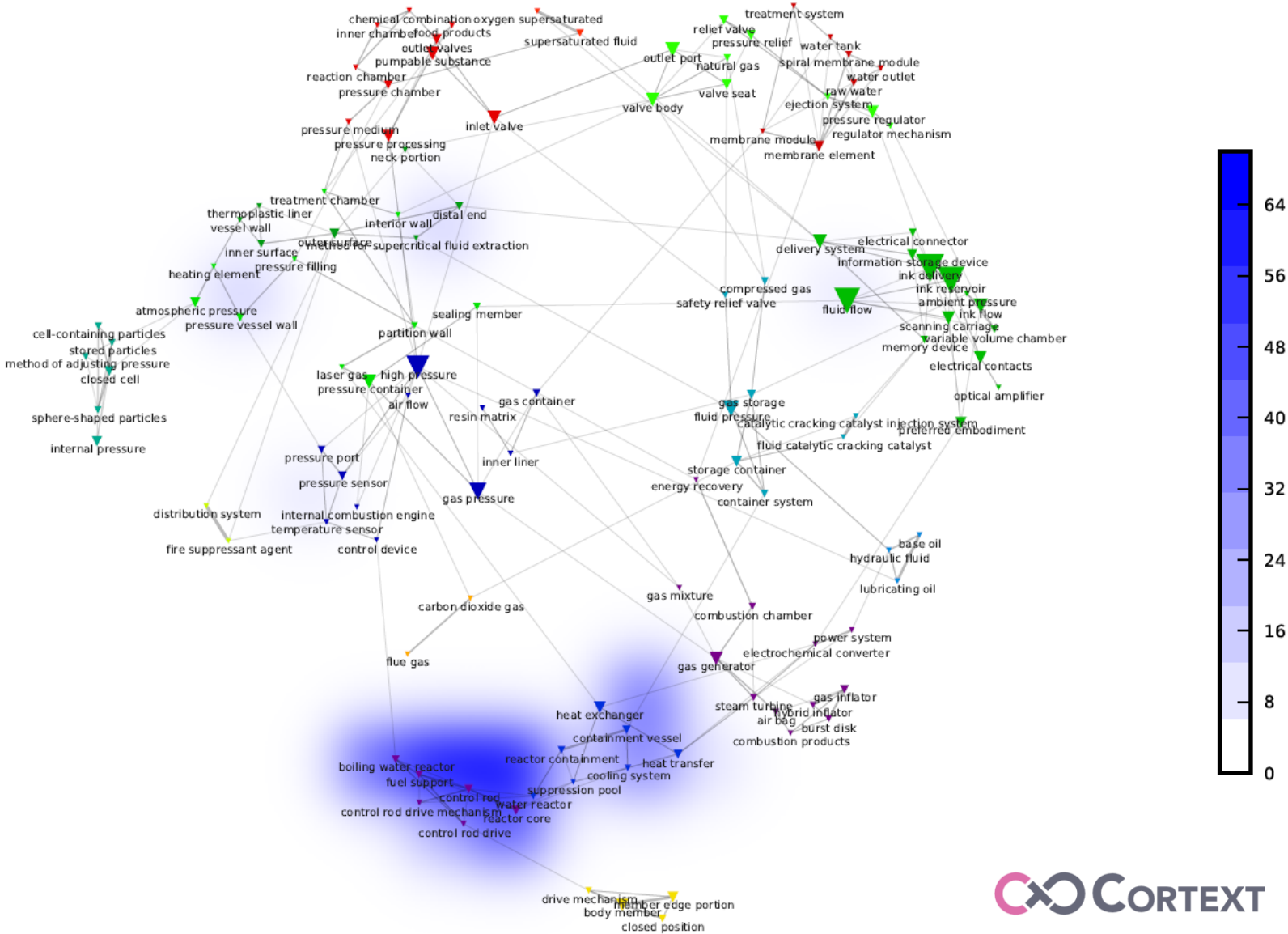


Publications scientifiques  
allemandes, Pressure vessel,  
entre 1991 et 2012

Spécialisations :

- Températures et matériaux
- Radiations

## GENERAL ELECTRIC, 1991-2011

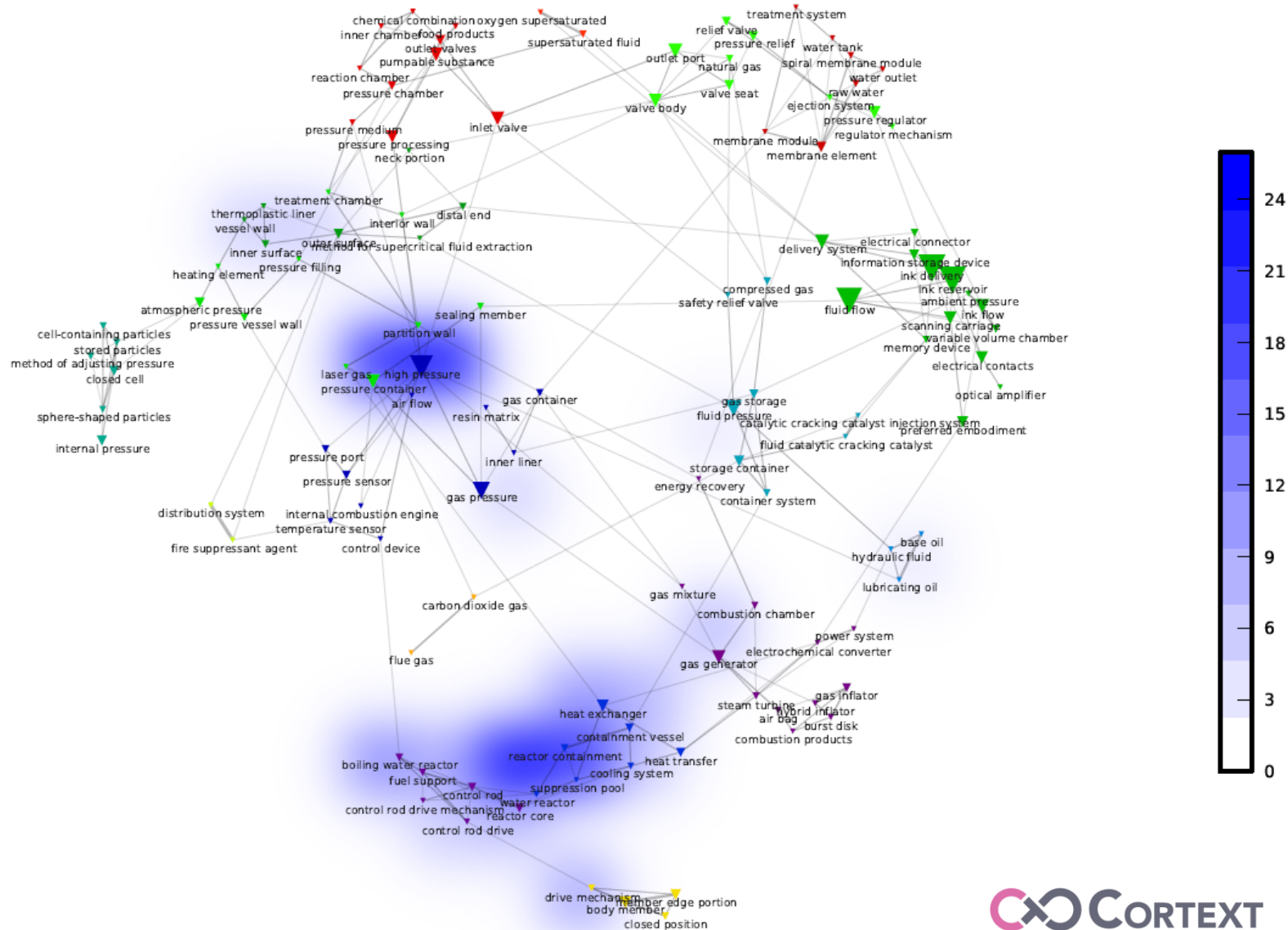


Dépôts prioritaires de General Electric,  
Pressure vessel, entre 1991 et 2012

Spécialisations :

- Cœur du réacteur, et contrôle

## HITACHI, 1991-2011



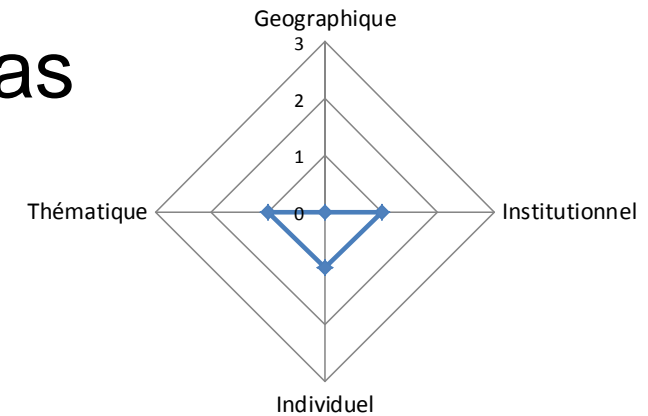
Dépôts prioritaires d'Hitachi, Pressure vessel, entre 1991 et 2012

Spécialisations :

- Système de refroidissement
- Container haut pression

# Co-activité (recherches et inventions) dans un champ scientifique et technologique : le cas de la catalyse hétérogène

Catalyse Hétérogène



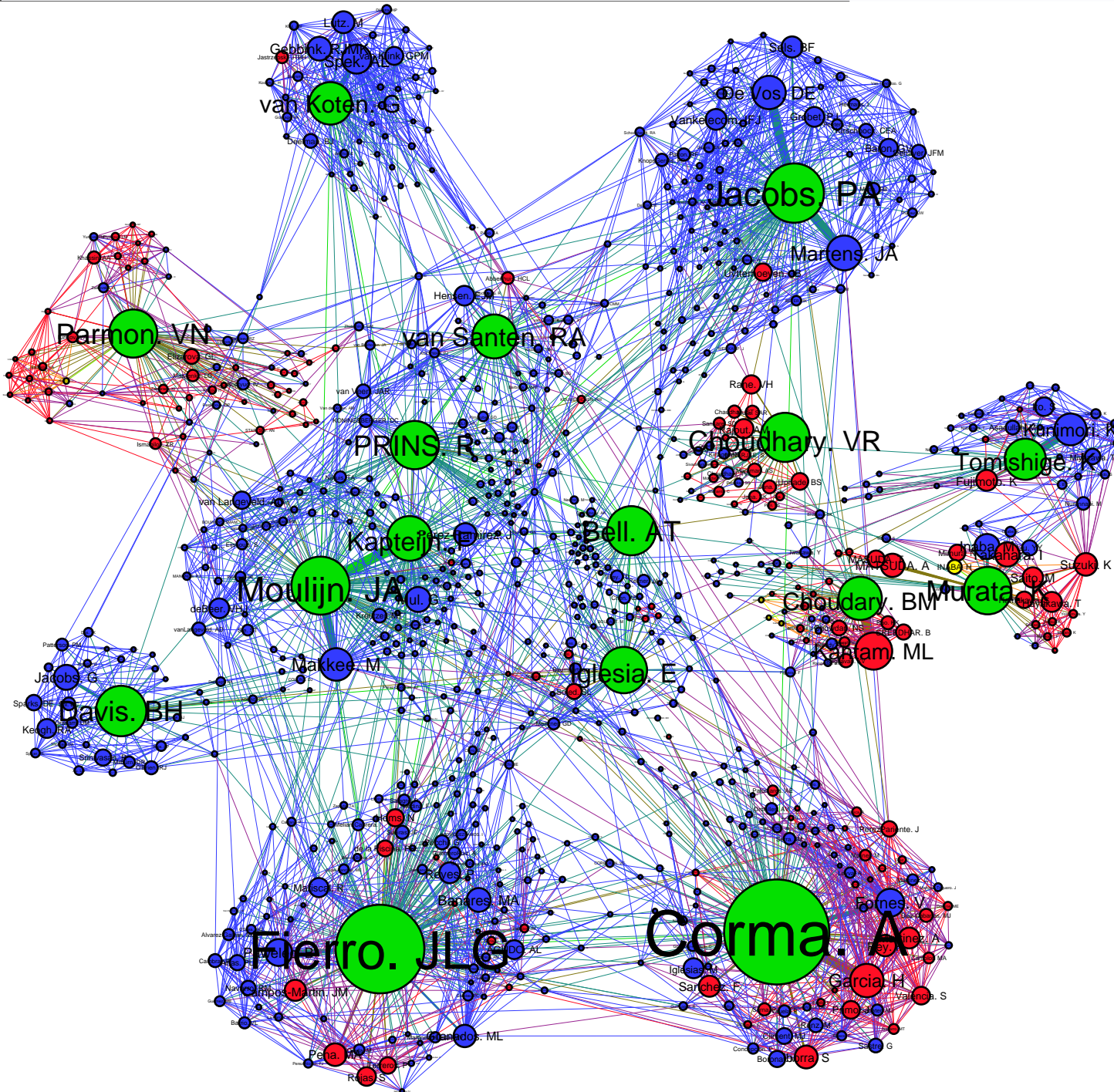
## Délinéation

- Identification d'un top16 d'auteurs centraux dans le champs scientifique de la catalyse hétérogène par enquêtes auprès d'experts.
- Publications et brevets associés

## Éléments principaux de méthode

- Réseau de collaborations de données hétérogènes





Analyse temporelle

Propension dans le temps à la  
co-activité en fonction de la  
structure du réseau de  
collaborateurs immédiats  
(brevets>publications)

## **Références mentionnées**

*Sémiologie graphique – Les diagrammes – les Réseaux – les cartes*, Jacques Bertin, 2005 4ème ed., ed EHESS.

*Atlas of science - Visualizing What We Know*, Katy Borner, 2010, The MIT Press.

*Mining, Mapping, and Acceleration Science and Technology*, Katy Borner, 2012, Sciences Po, paris

*Scholarly networks: networks, approaches, and applications*, Erjia Yan, 2012, Indiana University

*Information Graphics*, Sandra Rendgen, 2012, Taschen

*Science overlay maps: a new tool for research policy and library management*, Ismael Rafols, Alan L. Porter, Loet Leydesdorff, 2010, ASIS&T

*Industrial R&D Investment Scoreboard*, European Commission, Institut for Prospective Technological, 2008

*Data production methods for harmonized patent statistics: Patentee name harmonization*, Tom Magerman, Joris Grouwels,

Xiaoyan Song, Bart Van Looy (2009)

*Harmonizing Harmonized Patentee Names : an exploratory assessment of top patentees*, Bert Peeters, Xiaoyan Song, Julie Callaert, Grouwels Joris, Bart Van Looy (2010)

*A density-based algorithm for discovering clusters in large spatial databases with noise*, M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, in Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining, 1996, pp. 226–231.

*Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking*, Andrei Mogoutov, Bernard Kahane, Research Policy, 2007

*Chameleon Hierarchical clustering using dynamic modeling*, Karypis, Han, Kumar, IEEE Computer, 1999

*Concept of a Technology Classification for Country Comparisons*, Ulrich Schmoch, World Intellectual Property Organisation (WIPO), Fraunhofer, 2008

## **Références web**

*History of books*, Wikipedia , visité le 06/02/2010, [http://en.wikipedia.org/wiki/History\\_of\\_books](http://en.wikipedia.org/wiki/History_of_books)

*Data, data everywhere*, The economist, 25/02/2010, [www.economist.com/node/15557443](http://www.economist.com/node/15557443)

*Crunching the numbers*, The economist, 19/05/2012, [www.economist.com/node/21554743](http://www.economist.com/node/21554743)

## **Ressources web**

*Sciences, technologies et visualisations*, [www.sciences-technologies.eu](http://www.sciences-technologies.eu)

*Corporate Invention Board*, [www.corporateinventionboard.eu](http://www.corporateinventionboard.eu)

*CorText Manager*, <http://manager.cortext.net>

*Plateforme CorText*, [www.cortext.fr](http://www.cortext.fr)



Merci de votre attention